DOCUMENT RESUME

ED 421 528 TM 028 854

AUTHOR Bridgeman, Brent; Cooper, Peter
TITLE Comparability of Scores on Word-Processed and Handwritten Essays on the Graduate Management Admissions Test.
PUB DATE 1998-04-00
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Entrance Examinations; *College Students; Comparative Analysis; *Essay Tests; Graduate Students; *Handwriting; Higher Education; Interrater Reliability; Scores; Testing Problems; *Word Processing; Writing (Composition); Writing Evaluation; Writing Skills
IDENTIFIERS *Graduate Management Admission Test

ABSTRACT

Essays for the Graduate Management Admissions Test must be written with a word processor (except in some foreign countries). The test sponsors, the Graduate Management Admissions Council, believed that this is fair because some word processing skill is a prerequisite for advanced management education. Furthermore, it might also be unfair to require students who routinely use word processors to shift to paper and pencil just for a testing situation. The current study addressed the comparability of scores from handwritten and word-processed essays using a sample of 3,470 examinees who had written essays in both formats. Both the computer and paper-and-pencil versions contained two 30-minute essays questions, one asking for an analysis of an issue and the other analyzing the reasoning of a presented argument. Results indicate that scores were higher on the handwritten essays than on the word-processed essays, and that this difference did not interact with gender, ethnic, or English-as-a-Second-Language group classifications. Differences between scores for handwritten and word-processed essays were smallest for the most experienced computer users, but even examinees who reported using a word processor more than two times a week had higher scores on their handwritten essays than on their word-processed essays. Other findings indicated that reader reliability was higher for the word-processed essays, and that in either format there were substantial practice effects, with the scores on the second essay about 0.4 standard deviation units higher than scores on the first essay. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 421 528

Comparability of Scores on Word-Processed and Handwritten Essays on the Graduate Management Admissions Test

Brent Bridgeman and Peter Cooper

Educational Testing Service, Princeton, NJ

Brent Bridgeman

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM028854

Paper presented at the annual meeting of the American Educational Research Association, San Diego, April, 1998.

Abstract

Essays for the Graduate Management Admissions Test must be written with a word processor (except by examinees in some foreign countries that do not have access to computer testing centers). Although forcing all students to use a word-processor may seem to be unfair, the test sponsors, the Graduate Management Admissions Council, believed that some word-processing skill was a reasonable prerequisite for advanced management education. Furthermore, it might be equally unfair to require students who routinely use word processors to shift to paper and pencil just for a testing situation. The current study addressed the question of the comparability of scores from handwritten and word-processed essays using a sample of 3470 examinees who had written essays in both formats. Both the computer and paper-and-pencil versions contained two 30-minute essay questions, one of the two essay questions in each version required the student to write an analysis of an issue and the other question gave an argument and asked the student to write an essay analyzing the reasoning of this argument.

Results indicated that scores were higher on the handwritten essays than on the word-processed essays, and that this difference did not interact with gender, ethnic, or English as a Second Language group classifications. Differences between scores for handwritten and word-processed essays were smallest for the most experienced computer users, but even examinees who reported using a word processor more than two times a week had higher scores on their handwritten essays than on their word-processed essays. Other findings indicated that reader reliability was higher for the word-processed essays, and that in either format there were substantial practice effects, with scores on the second essay about .4 SD units higher than scores on the first essay.

Comparability of Scores on Word-Processed
and Handwritten Essays on the Graduate Management
Admissions Test

The use of word processors is ubiquitous on college campuses. Many students have come to rely on word processors for their college writing assignments. Thus, it would seem to be reasonable to assess the writing skills of college students with essays that were produced on word processors. Indeed, as of October, 1997, the essays for the Graduate Management Admissions Test must be written with a word processor at a computerized testing center (except for examinees in some foreign countries that do not have access to computer testing centers). Although it is assumed that candidates for graduate management programs should have some word-processing skills, some fairness concerns with this requirement remain. Although forcing all students to use a word-processor may seem unfair, it might be equally unfair to require students who routinely use word processors to shift to paper and pencil just for a testing situation.

A comprehensive review of the effects of word processors on the quality of students' writing has shown mixed results (Cochran-Smith, 1991). Many of the studies reviewed focus on the role of the word processor in helping students make revisions over several drafts and may not generalize to a testing situation in which only 30 minutes are allowed from first reading of the question to final essay. Furthermore, findings from students in elementary and secondary schools who have relatively little word-processing experience may not generalize to experienced word-processor users in college. A study of college students found that scores assigned to word-processed essays were fairly comparable to scores assigned to handwritten essays produced by the same students (Powers, Fowles, Farnum, & Ramsey, 1992), with a slight advantage in producing essays on the computer offset by a tendency to grade handwritten essays more leniently. The sample of students in the Powers et al. study was very small (32), so separate analyses by subgroups were not feasible. The current study was designed to assess the comparability of word-processed and handwritten GMAT essays for different gender, ethnic, and language fluency groups, and for examinees with differing amounts of word-processing experience.

Methods and Data Source

A random sample of students who registered to take the regular paper-and-pencil administration of the GMAT in October 1996 were invited to also take the new computerized version of the GMAT in October, including using the computer to word process the essays. A random half of the sample was invited to take the computerized test first with the other half taking the paper-and-pencil version first. The computerized test was free, and volunteers were told that their scores on the computer test would replace the scores on the paper-and-pencil test if and only if they were higher. Thus, students had nothing to lose, and possibly higher scores to gain, by taking the computerized version. Students identified their level of word processing experience on a posttest questionnaire. Categories on frequency of word-processor use ranged on a five point scale from never to more than two times per week.

Both the computer and paper-and-pencil versions contained two 30-minute essay questions, one of the two essay questions in each version required the student to write an analysis of an issue and the other question gave an argument and asked the student to write an essay analyzing this argument. For the computer-delivered tests, there were 12 issue topics and 12 argument topics. The computer randomly selected one topic of each type for each person. Order was counterbalanced such that an issue essay was first for half of the sample and an argument essay was first for the other half. For the handwritten essays that were part of the regular GMAT October administration, there was only one argument topic, there were two issue topics (one for the Americas and one for the rest of the world). All students responded to the argument topic first.

All essays were read by two readers with a third reader used if the scores differed by more than one point. Each reader assigned a score of 1 to 6 on a holistic scale. The scores from the readers were averaged. Readers for the word-processed essays were a subset of the readers for the handwritten essays.

Results

Usable data were obtained from 3470 examinees who completed the test in both formats. Samples were smaller for some analyses; for example, only U. S. citizens are asked to provide ethnic group and some groups (e.g., American Indians) did not have sufficient numbers to be analyzed separately, resulting in a sample of 2453 examinees in four major ethnic groupings (African American, Asian American, Hispanic, and White). A separate analysis, that included non-U.S. citizens, compared the 2337 examinees whose best language was English with the 775 examinees who were most fluent in a language other than English.

Scores from both topics in the paper-and-pencil mode were added to make a handwritten essay total, and a word-processed essay total was similarly constructed. The word-processed essay total was subtracted from the handwritten essay total to form a difference score with positive values indicating higher scores on the handwritten essay. As shown in Table 1, values for all subgroups were positive, with relatively little variation among gender and ethnic subgroups. A 2 (genders) x 3 (ethnic groups) by 5 (levels of word-processing experience) ANOVA indicated a significant effect ($p = .04$) only for word-processing experience. A similar analysis contrasting the 775 examinees who were most fluent in a language other than English with the 2337 fluent English speakers produced similar nonsignificant results for fluency but a significant experience effect.

Rater reliability was estimated from the correlation between the two raters adjusted by the Spearman-Brown formula. Rater reliability was the same for issue essays and for analysis of an argument essays, but it was higher for the word-processed essays than for handwritten essays (.87 versus .80). This probably reflects the greater standardization in the word-processed essays in which raters cannot attend to differences in handwriting or overall neatness. Apparently because of this higher reliability, scores on the word-processed essays were more highly correlated with scores on the verbal scale than were scores from the handwritten essays (.60 versus .54).

There were significant practice effects both across formats and within the word-processed format. For examinees who took the computer test first, scores were .43 points higher on the handwritten tests (SD = .72), and for students who took the handwritten test first, scores were only .16 higher on the handwritten test (SD = .69). Assuming that practice effects were constant across modes, these numbers are consistent with a practice effect of .13 points and a mode effect of .29 points. As indicated in Table 2, for the word-processed essays, there was a substantial gain from the first topic to the second, regardless of which topic type was first. For the handwritten essays there was also a substantial gain; mean on the argument essay (which was always first in the handwritten administration) was 3.84 (SD = .96) and the mean on the issue essay was 4.19 (SD = .95), for a gain of .35 points on the 1 to 6 scale.

Educational Importance

Moving from handwritten to word-processed essay assessments would appear to have positive benefits in terms of enhanced reliability. Furthermore, this switch would not appear to disadvantage gender, ethnic, or language minority subgroups relative to handwritten assessments. However, caution is needed because of the high level of word-processing experience in this sample of examinees bound for graduate management training, and the indication that students with less experience may have relatively more difficulty with word-processed essays. The data on practice effects suggest that students would be well advised to practice writing essays on a word processor, with GMAT-type topics and timing conditions, before attempting to take the actual examination.

References

Cochran-Smith, M. (1991). Word processing and writing in elementary classrooms: A critical review of related literature. *Review of Educational Research, 61*, 107-155.

Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1992). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement, 31*, 220-233.

Table 1

Paper Essay Total Score Minus Computer Essay Total Score by Gender, Ethnicity, and Word Processing Experience Level

		Demographic Group								
		Asian American		African American		Hispanic		White		Total
WP Experience	Statistic	M	F	M	F	M	F	M	F	
	N	4	2	2	2	1	1	14	6	32
1	M	63	50	1.0	0.0	0.0	50	82	67	67
	SD	63	1.4	71	0.0	--	--	80	88	76
	N	8	11	7	4	1	3	45	34	114
2	M	63	64	86	63	50	50	51	37	51
	SD	95	45	69	1.0	--	1.0	66	75	71
	N	27	14	14	16	6	7	107	59	350
3	M	37	29	43	63	08	29	33	40	37
	SD	70	83	68	85	49	1.0	68	71	71
	N	27	21	9	18	11	8	161	92	348
4	M	33	26	67	31	45	06	30	42	34
	SD	72	64	90	86	69	42	74	65	71
	N	117	88	49	83	55	52	730	537	1712
5	M	18	21	11	36	23	49	19	28	23
	SD	67	55	55	72	71	72	44	71	71
	N	183	136	81	123	74	71	1057	728	2453
Total	M	25	26	31	39	25	42	24	31	28
	SD	73	61	68	76	68	74	74	71	72
% In Experience Level 1 or 2		7%	10%	11%	5%	3%	6%	6%	5%	6%

Note.--Word-processing experience levels: 1 = never; 2 = <once a month; 3 = between once a week and once a month; 4 = 1 or 2 times a week; 5 = more than 2 times a week.

Table 2

Gain from First to Second Word-Processed Essay

Topic and Order	Mean	SD	Gain
Argument First	3.46	1.06	
Issue Second	3.91	1.06	.45
Issue First	3.59	1.06	
Argument Second	3.89	1.07	.30